UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

MODÉLISATION DES CHOIX DE RESIDENCES

par

Alain Lapointe et Jacques Desrosiers

Ce rapport de recherche a été publié grâce à une subvention du Fonds F.C.A.C. pour l'aide et le soutien à la recherche.

Université de Montréal

Centre de recherche sur les transports - Publication #373
Ecole des Hautes Etudes Commerciales - Cahier IEA-84-01

Juillet 1984

Dépôt légal - 4e trimestre 1984

RESUME

L'objectif de l'étude est d'évaluer l'impact sur les choix résidentiels de certains changements structurels tels les modifications dans la composition de la population ou encore du stock de logements. Le modèle de simulation proposé s'articule autour de trois modules spécifiques. Dans un premier module, on estime les probabilités individuelles de choix à l'aide du modèle logit multinomial. Un deuxième module sert à agréger les ménages par classes homogènes selon le revenu, l'âge du chef et la taille des ménages. On obtient ainsi les probabilités agrégées ex-ante. Dans un dernier module, les probabilités ex-ante sont modifiées à l'aide d'un modèle de programmation non linéaire basé sur la théorie de l'information. On obtient ainsi des probabilités ex-post qui respectent la composition du stock de logements.

Le modèle est estimé à partir d'un échantillon de ménages de la région métropolitaine de Montréal. Deux simulations sont présentées. Une première portant sur un scénario de vieillissement de la population et une seconde basée sur un scénario de diminution de la taille des ménages. Ces deux simulations ne représentent toutefois qu'un pâle reflet des possibilités d'utilisation de ce type de modèle.

La plupart des modèles développés jusqu'à maintenant dans le domaine urbain, que ce soit les modèles économétriques ou encore les modèles déterministes de simulation, n'ont généralement pas fourni les résultats qu'on en attendait. Les tentatives les plus ambitieuses dans ce sens furent certainement les modèles du URBAN INSTITUTE (7) et du NBER (12). L'expérience avec ces modèles a permis de mettre en évidence les difficultés de combiner une approche théorique rigoureuse avec une vérification empirique qui présente les mêmes qualités. Ces expériences ont contribué à entretenir un certain scepticisme envers les modèles d'une telle complexité.

Au même moment où se développait une certaine désillusion envers ces modèles, des modifications structurelles importantes s'opéraient sur le marché du logement. L'évolution démographique, caractérisée par une baisse de la natalité et le vieillissement de la population, modifie considérablement la composition de la demande de logement de même que l'attrait relatif de l'accessibilité à l'emploi par rapport à l'espace. Ces modifications structurelles majeures couplées à l'augmentation des prix relatifs du transport ont affecté considérablement les choix résidentiels tant du point de vue de la localisation à l'intérieur de l'espace urbain que du point de vue du type d'habitat. Ces développements reposent avec toute sa pertinence la question de l'utilité des modèles de simulation, soit des modèles permettant d'évaluer dans un cadre cohérent l'impact de ces changements.

Le développement récent des modèles de choix probabilistes a permis de faire le lien entre les fondements théoriques de l'analyse des comportements individuels et de l'observation empirique de ces comportements où se confondent des composantes stochastiques et déterministes. La présente étude a pour objectif de développer un modèle de simulation des changements structurels qui s'appuie sur ces nouveaux développements. La partie originale du modèle de simulation proposée est certainement de combiner ce type de modèles probabilistes avec un modèle de programmation mathématique aux fins de la simulation.

Nous présentons donc dans un premier temps l'architecture générale du modèle de simulation. Nous reproduisons ensuite les principaux résultats de l'estimation du modèle pour la Région Métropolitaine de Montréal sur la base d'une enquête effectuée par l'INRS-URBANISATION en 1978 (6). La dernière partie est consacrée à la présentation de deux simulations servant à illustrer l'utilisation du modèle.

1. L'ARCHITECTURE GENERALE DU MODELE

Le modèle de simulation proposé s'articule autour de trois modules spécifiques (figure 1). Dans le premier module, connaissant l'ensemble des types de logement et les variables socio-économiques influençant les choix des ménages, les probabilités individuelles de choix sont estimées à l'aide du modèle logit multinomial.

En second lieu, à des fins prévisionnelles, nous agrégeons les ménages par classes homogènes selon le revenu, l'âge du chef et la taille des ménages. On obtient ainsi des probabilités agrégées ex-ante qui sont par la suite ajustées pour respecter la composition du stock de logements dans l'échantillon.

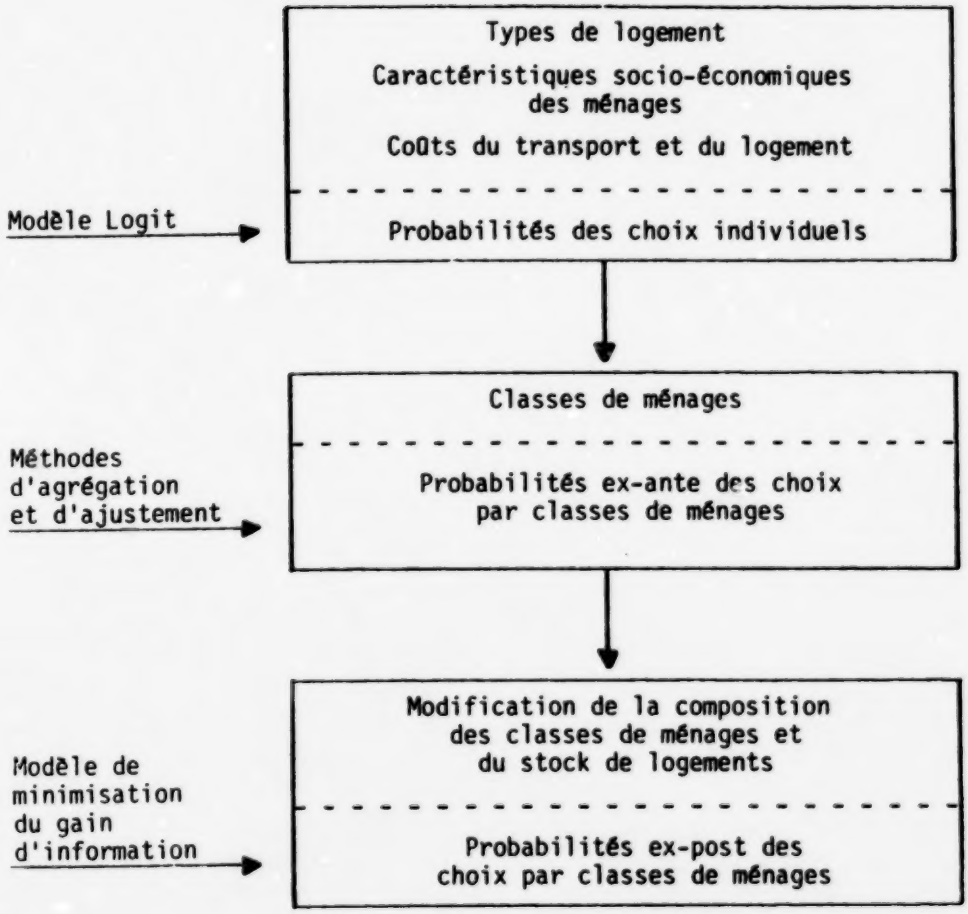
Finalement, le troisième module simule l'effet sur les probabilités de choix agrégées de scénarios de changements de la composition des classes de ménages et du stock de logements. Les nouvelles probabilités sont évaluées à l'aide d'un modèle de programmation non linéaire basée sur la théorie de l'information qui transforme les probabilités ex-ante en probabilités ex-post suite à des modifications sur l'offre et la demande.

1.1 La modélisation des probabilités ex-ante

Supposons un échantillon de h = 1,...,H ménages individuels où chacun de ces ménages est confronté à un ensemble de J alternatives en ce qui concerne son choix de logement. Chaque ménage h évalue chaque type de logement j selon le niveau d'utilité U_j^h qu'il lui procure. Cette fonction d'utilité peut s'exprimer par

FIGURE 1

MODELE DE SIMULATION



1) $U_j^h = \alpha_{0j} + \sum_k \alpha_k X_{jk}^h + \varepsilon_j^h$

où $\underline{\alpha} = (\alpha_{01}, \alpha_{02} \cdots \alpha_{0J}, \alpha_1, \ldots, \alpha_k)$ sont les coefficients de la fonction d'utilité commune à tous les ménages dans la population, X_{jk}^h est la valeur du kème attribut pour le type de logement j et le ménage h et ε_j^h est la composante aléatoire. Les coefficients α_{0j} sont des constantes spécifiques qui permettent de capter l'influence moyenne de la partie inexpliquée de l'utilité associée à chaque alternative, et $V_j^h = \alpha_{0j} + \sum_k \alpha_k X_{jk}^h$ représente la partie systématique de l'utilité appelée l'utilité stricte.

Le ménage choisit l'alternative pour laquelle il dérive la plus grande utilité de telle sorte que la probabilité que le ménage h choisisse le type de logement j est définie par:

2) $P_j^h = \text{Prob}\ [U_j^h \geq U_s^h\ ;\ \text{pour tout } s \neq j]$.

Si les composantes aléatoires ε_j^h sont distribuées de façon identique et indépendante selon une distribution de Weibull, alors la probabilité que le ménage h choississe le type de logement j devient:

3) $$P_j^h = \frac{\exp\ (\alpha_{0j} + \sum_k \alpha_k X_{jk}^h)}{\sum_s \exp\ (\alpha_{0s} + \sum_k \alpha_k X_{sk}^h)} = \frac{\exp\ (V_j^h)}{\sum_s \exp\ (V_s^h)}$$.

Il s'agit du modèle logit multinominal développé et appliqué par Domencich et McFadden (9), McFadden (15) et Anas (1).

Pour estimer ce modèle, on doit maximiser la fonction de vraisemblance par rapport au vecteur des coefficients $\underline{\alpha}$ soit:

$$4) \quad \max_{\underline{\alpha}} \log L = \sum_h \sum_j \delta_j^h \log P_j^h (\underline{\alpha})$$

où $\delta_j^h = \begin{cases} 1 \text{ si le ménage h choisit l'alternative j} \\ 0 \text{ autrement} \end{cases}$

Les coefficients $\underline{\alpha}$ peuvent être estimés de façon unique à l'exception d'une constante spécifique disons α_{0J}. Ainsi, les différences $\alpha_{0j} - \alpha_{0J}$ sont uniquement définies.

Ce modèle est dérivé de postulats de comportements de la part des ménages individuels. Anas (2) a montré que les probabilités P_j^h peuvent être également obtenues à partir d'hypothèses beaucoup moins restrictives, soit par la maximisation de la mesure d'information de Shannon I(P) (voir Renyi (18)) définie sur la distribution de probabilité P, sujet aux contraintes décrivant certaines propriétés du système de choix. Le modèle de programmation non linéaire est le suivant:

$$5) \quad \max I(P) = \max \sum_h \sum_j P_j^h \log \left(\frac{1}{P_j^h} \right)$$

$$6) \quad \equiv \min \sum_h \sum_j P_j^h \log P_j^h$$

$$7) \quad \text{sujet à} \quad \sum_j P_j^h = 1 \qquad h=1,\ldots,H$$

$$8) \quad \sum_h P_j^h = M_j \qquad j=1,\ldots,J$$

$$9) \quad \sum_h \sum_j P_j^h X_{jk}^h = X_k \qquad k=1,\ldots,K$$

où $M_j = \sum_h \delta_j^h$ est le nombre d'unités du type de logement j choisies par l'ensemble des ménages et $X_k = \sum_h \sum_j \alpha_j^h X_{jk}^h$ est la valeur agrégée du kème attribut. Le problème 6)-9) est la minimisation d'une fonction non linéaire strictement convexe sujet à des contraintes linéaires. Les conditions d'optimalité de Kuhn-Tucker sont nécessaires et suffisantes et les valeurs optimales uniques P_j^h solution de ce problème sont également données par l'équation 3). Dans cette formulation, le vecteur $\underline{\alpha}$ correspond aux variables duales associées aux contraintes 8) et 9).

1.2 La procédure d'agrégation: les probabilités agrégées ex-ante

Une fois que le modèle logit multinomial a été estimé, pour les fins de la simulation, les ménages sont regroupés dans des classes homogènes. L'agrégation doit être faite de telle sorte que l'ensemble des choix soit le même à l'intérieur d'un même groupe et que les variables explicatives aient approximativement la même valeur à l'intérieur de chaque groupe de telle sorte qu'il est raisonnable d'utiliser la valeur moyenne de ces variables.

On peut comprendre les mécanismes d'agrégation en considérant la formulation exacte suivante pour les probabilités agrégées ex-ante. Pour l'ensemble des ménages de la classe i, noté C_i $(i = 1,\ldots,I)$, soit $N_i = \sum_{h\in C_i} \sum_j \delta_j^h$ le nombre de ménages et $\bar{X}_{ijk} = \sum_{h\in C_i} X_{jk}^h / N_i$ la valeur moyenne du kème attribut pour le

type de logement j. On cherche les probabilités agrégées de telle sorte que $P_j^h = P_{ij}$ pour tout $h \in C_i$. En se référant à la formulation du modèle présenté aux équations 6) à 9), on obtient par substitution:

$$10) \quad \text{Min} \sum_i \sum_j N_i P_{ij} \log P_{ij}$$

$$11) \quad \text{Sujet à} \sum_j P_{ij} = 1 \qquad (i=1,\ldots,I)$$

$$12) \quad \sum_i N_i P_{ij} = M_j \qquad (j=1,\ldots,J)$$

$$13) \quad \sum_i \sum_j N_i P_{ij} \bar{X}_{ijk} = X_k \qquad (k=1,\ldots,K).$$

La probabilité pour un ménage appartenant à la classe i de choisir le type de logement j peut alors s'exprimer par:

$$14) \quad P_{ij} = \frac{\exp\left(\lambda_{0j} + \sum_k \lambda_k \bar{X}_{ijk}\right)}{\sum_s \exp\left(\lambda_{0s} + \sum_k \lambda_k \bar{X}_{isk}\right)}$$

où λ_{0j}, (j=1,...,J) et λ_k, (k=1,...,K) sont les variables duales correspondant aux contraintes 12) et 13) du modèle précédent. Ce modèle évaluant les probabilités agrégées ex-ante peut être résolu par des méthodes de balancement (voir Censor et Herman (4) et Elfving (10)) qui sont des méthodes non linéaires de type Gauss-Seidel. Nous leur avons préféré la procédure plus classique en deux phases, soit le calcul des probabilités désagrégées par le modèle logit multinomial (dont les codes offrent beaucoup de flexibilité sur la méthodologie d'estimation) suivie d'une procédure d'agrégation.

Si les probabilités P_{ij} sont évaluées approximativement en utilisant les coefficients estimés $\underline{\alpha}$ du modèle désagrégé à la place des coefficients $\underline{\lambda}$ de l'équation 14), alors

$$15) \quad P_{ij} = \frac{\exp\left(\alpha_{0j} + \sum_k \alpha_k \bar{X}_{ijk}\right)}{\sum_s \exp\left(\alpha_{0s} + \sum_k \alpha_k \bar{X}_{isk}\right)}$$

et ces probabilités ne respectent pas nécessairement les contraintes 12) sur le nombre d'unités choisies dans l'échantillon du type de logement j, ni les contraintes 13) décrivant les propriétés sur les attributs.

Par contre, si les probabilités P_{ij} sont évaluées approximativement en prenant la moyenne des P_j^h sur chaque classe de ménage i, soit:

$$16) \quad P_{ij} = \sum_{h \in C_i} P_j^h / N_i$$

alors les contraintes 12) sont respectées, les contraintes 13) ne le sont pas et les probabilités P_{ij} ne peuvent plus s'exprimer sous la forme d'un modèle logit et ne peuvent pas être utilisées comme outil de simulation à partir des valeurs moyennes des attributs par classe. Ainsi, la première méthode d'agrégation semble préférable mais elle ne reproduit pas les résultats globaux de l'échantillon. Nous avons complété cette méthode d'agrégation qui utilise les coefficients $\underline{\alpha}$ à la place des coefficients $\underline{\lambda}$ par un modèle d'optimisation non linéaire de telle sorte que les probabilités agrégées sont ajustées selon le nombre d'unités choisies du stock de logements dans l'échantillon. Ce

modèle est présenté à la section suivante et il est décrit dans le contexte de la détermination des probabilités agrégées ex-post suite à des changements structurels.

1.3 Modélisation des probabilités agrégées ex-post

Sur la base des probabilités de choix ex-ante P_{ij}, on veut obtenir la distribution des probabilités ex-post Q_{ij} consécutive à des changements structurels tels la croissance démographique, le vieillissement de la population ou la modification dé la composition du stock. Pour fins de simulation, on suppose que le nombre de ménages N_i dans la classe i et que le stock disponible S_j pour le type de logement j sont déterminés de façon exogène. On suppose également que le stock total est suffisant, soit:

$$17) \quad \sum_j S_j \geq \sum_i N_i = N$$

Les probabilités ex-post Q_{ij} sont obtenues en minimisant la mesure du gain d'information du Shannon I(Q,P) (voir Hobson (11) et Renyi (18)) sujet aux contraintes sur le stock, soit:

$$18) \quad \text{Min } I(Q,P) = \text{Min} \sum_i \sum_j N_i \, Q_{ij} \log \left(\frac{Q_{ij}}{P_{ij}} \right)$$

$$19) \quad \text{sujet à} \sum_j Q_{ij} = 1 \qquad (i=1,\ldots,I)$$

$$20) \quad \sum_i N_i \, Q_{ij} \leq S_j \qquad (j=1,\ldots,J).$$

L'idée d'utiliser la théorie de l'information dans ce contexte revient à Snickars et Weibull (19). Pour un exposé du champs d'application de ce modèle, le lecteur peut également se référer à Los (14).

La fonction $I(Q,P)$ est considérée comme une mesure de l'information contenue dans le message qui transforme P en Q. Elle caractérise la proximité entre deux distributions de probabilités. Pour une meilleure compréhension du modèle, on peut l'exprimer en termes des demandes ex-ante $N_{ij} = N_i P_{ij}$ et des demandes inconnues ex-post $Z_{ij} = N_i Q_{ij}$. Le modèle prend alors la forme suivante:

$$21) \quad \text{Min} \sum_i \sum_j Z_{ij} \log\left(\frac{Z_{ij}}{N_{ij}}\right)$$

$$22) \quad \text{sujet à} \sum_j Z_{ij} = N_i \qquad (i=1,\ldots,I)$$

$$23) \quad \sum_i Z_{ij} \leq S_j \qquad (j=1,\ldots,J).$$

Ce dernier problème minimise une fonction strictement convexe, sujet à des contraintes linéaires décrivant un problème de transport. La solution primale unique nous est donnée par

$$24) \quad Z_{ij} = N_{ij} \exp(-1 + u_i + v_j)$$

ou de façon équivalente pour les probabilités ex-post,

$$25) \quad Q_{ij} = P_{ij} \exp(-1 + u_i + v_j)$$

où u_i et v_j sont les variables duales associées aux contraintes 22) et 23). Puisque $\sum_j Q_{ij} = 1$ on obtient:

$$26) \quad Q_{ij} = \frac{P_{ij} \exp(v_j)}{\sum_s P_{is} \exp(v_s)} .$$

On remarquera que les variables duales v_j sont définies à une constante additive près (équation 26) et que l'ajout d'une constante aux v_j nécessite la soustraction de cette même constante aux variables duales u_i (équation 24). Pour un couple (i,j), si $u_i + v_j = 1$, alors les probabilités Q_{ij} et P_{ij} sont identiques, alors qu'il y a croissance ou décroissance des probabilités ex-post selon que $u_i + v_j > 1$ ou $u_i + v_j < 1$ (équation 25).

Finalement, lorsque les probabilités ex-ante P_{ij} sont exprimées à l'aide d'un modèle logit multinomial avec l'utilité stricte donnée par une certaine fonction V_{ij}, alors les probabilités ex-post Q_{ij} s'expriment sous la même forme, soit:

$$27) \quad Q_{ij} = \frac{\exp (v_j + V_{ij})}{\sum_s \exp (v_s + V_{is})} .$$

Comme nous l'avons déjà mentionné, ce modèle a été appliqué aux probabilités agrégées ex-ante P_{ij} de telle sorte que les probabilités ajustées respectent les contraintes de stock dans l'échantillon. Le modèle devient, en terme des demandes,

$$28) \quad \text{Min} \sum_i \sum_j Z_{ij} \log \left(\frac{Z_{ij}}{N_{ij}} \right)$$

29) sujet à $\sum_j Z_{ij} = N_i \qquad (i=1,\ldots,I)$

30) $\sum_i Z_{ij} = M_j \qquad (j=1,\ldots,J).$

Les probabilités ajustées Z_{ij}/N_i montrent alors une correction pour les constantes spécifiques α_{0j} par rapport à l'équation 15):

31) $$\frac{Z_{ij}}{N_i} = \frac{\exp\left(v_j + \alpha_{0j} + \sum_k \alpha_k \bar{X}_{ijk}\right)}{\sum_s \exp\left(v_s + \alpha_{0s} + \sum_k \alpha_k \bar{X}_{isk}\right)} \quad .$$

2. ESTIMATION DU MODELE

L'opérationalisation du modèle de simulation précédent nécessite que l'on définisse un ensemble de choix, que l'on estime les probabilités de choix individuels, que l'on adopte une procédure d'agrégation et d'ajustement afin d'obtenir les probabilités agrégées ex-ante et enfin que l'on détermine les probabilités agrégées ex-post suite à des scénarios de changements structurels. Nous allons présenter les principales étapes de cette démarche en se limitant à reproduire les résultats qui nous semblaient les meilleurs et les plus probants. Les estimations portent sur un échantillon de 1,357 ménages obtenu d'une enquête de l'INRS-URBANISATION réalisée en 1978. Pour de plus amples renseignements, le lecteur peut se référer à Desrosiers et Lapointe (8).

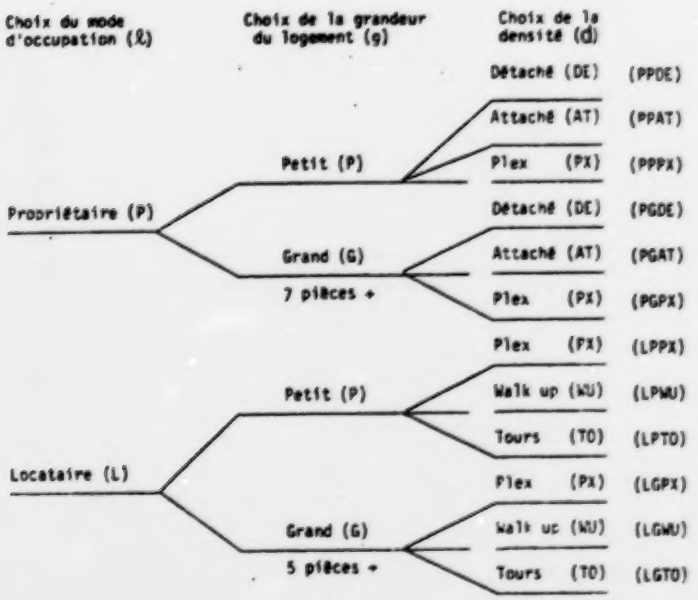
2.1 L'ensemble des choix et la fonction d'utilité stricte

Une des caractéristiques distinctes du bien logement est certainement son degré d'hétérogénéité. A la limite, il n'y a pas deux logements semblables si ce n'est qu'en raison de leur localisation différente. Aussi, pour définir l'ensemble des choix, on réfère aux principaux attributs du logement: le mode d'occupation, la grandeur du logement, le type de bâtiment et sa localisation. Ces diverses dimensions du bien logement permettent de caractériser assez bien l'ensemble des choix. Cependant, si on distingue 2 modes d'occupation, 2 catégories de grandeur et 3 types de densité, on a déjà 12 types de logement et la localisation n'a pas été prise en compte. C'est pourquoi l'aspect de la localisation n'est pas explicité directement au

niveau de l'ensemble des choix mais plutôt par le biais d'une des variables explicatives du modèle: le prix effectif du logement. (Cette approche revient à Quigley (17)). L'ensemble des choix retenus est présenté à la figure 2.

Figure 2

Définition des types de logements



La définition des attributs diffère selon le mode d'occupation. Par exemple, le grand logement propriétaire comprend 7 pièces et plus alors qu'il ne comprend que 5 pièces et plus s'il est locatif. Le plex réfère aux logements situés dans un immeuble de 5 logements ou moins dans lequel chaque logement possède sa propre adresse. Le walk-up réfère aux immeubles de 3 logements ou moins sans ascenseur.

On suppose que le ménage, dans son choix du logement, incorpore l'information sur la localisation optimale pour chaque type de logement. Cette information est contenue dans le prix effectif minimum, c'est-à-dire la dépense en logement et en transport minimale pour consommer un type de logement compte tenu du lieu d'emploi et du revenu du ménage. Par exemple, dans le cas du transport par automobile, on peut dériver le prix effectif minimum D_{jrd} de la façon suivante:

38) $$D_{jrd} = \min_{o} \left[R_{jo} + T_{rod}\right]$$

où $$T_{rod} = t_{od} (W_r + CO) + CF$$

R_{jo}: Les dépenses de logement de type j dans la zone de résidence o.

t_{od}: Le temps de déplacement en automobile de la zone de résidence o à la zone d'emploi d.

W_r: Le revenu du ménage.

CO: Les coûts d'opération d'une automobile standard.

CF: Les coûts fixes.

L'information sur les temps de déplacement provient de l'enquête origines/destinations du COTREM (5), les coûts d'opération et les coûts fixes sont publiés par l'Association Canadienne de l'Automobile (3). Une équation analogue peut être définie pour le transport public. On suppose que le ménage choisit le mode de transport le moins cher incluant le coût du temps. Le tableau 1 reproduit ces coûts moyens de transport selon le mode par catégorie de revenu et type de logement.

Tableau 1

COUTS MOYENS DE TRANSPORT SELON LE MODE
PAR CATEGORIE DE REVENU ET TYPE DE LOGEMENT(a)

	Moins de $15 000		$15 000-24 999		$25,000 et plus	
	Automobile	Transport public	Automobile	Transport public	Automobile	Transport public
Unifamilial détaché	$216,7	$150,4	$285,1	$342,6	$429,2	$578,9
Unifamilial attaché	182,8	149,7	269,5	292,9	375,1	501,3
Plex propriétaire	166,9	109,4	208,0	191,8	231,6	328,7
Plex locataire	157,6	85,6	189,5	191,3	249,1	370,1
Walk up	161,5	81,6	230,5	232,1	304,8	484,9
Tours	142,0	58,3	172,3	147,5	226,2	308,1

a) Les coûts sont calculés à partir de l'équation 38) pour l'automobile et d'une équation analogue pour le transport public.

On constate que le transport par automobile est plus efficace que le transport en commun et s'avère toujours moins dispendieux pour les catégories de revenu moyen et élevé.

Le deuxième élément important à définir préalablement à l'estimation est la spécification de la fonction d'utilité représentative tant du point de vue de la forme fonctionnelle que des variables à inclure. Sur ce dernier plan, nous distinguerons les variables spécifiques aux alternatives et les variables génériques. Le niveau de ces dernières varie d'une alternative à l'autre et elles apparaissent dans la fonction d'utilité représentative de chaque alternative. Par contre,

les variables spécifiques ont un lien directement identifiable aux alternatives.

Dans la présente étude, deux types de variables sont introduites; le prix minimal de chaque alternative et un ensemble de variables socio-économiques, soit le revenu, la taille du ménage et l'âge du chef. La première variable est introduite comme variable générique et les variables socio-économiques comme variables spécifiques. En effet, dans ce dernier cas, puisque le niveau de ces variables est indépendant des alternatives (par exemple, le revenu du ménage est le même quelle que soit l'alternative), leur influence s'annule à moins de les introduire comme des variables spécifiques. Elles prennent alors la valeur observée pour une ou plusieurs alternatives et zéro pour les autres. Elles exercent alors un pur effet de déplacement sur les choix. Par exemple, les ménages à revenus élevés ont une préférence plus grande, toutes choses étant égales par ailleurs, pour les grands logements détachés et détenus en propriété.

La fonction d'utilité stricte peut alors être définie de la façon suivante, pour chaque ménage h et type de logement j:

$$39) \quad V_j^h = \alpha_{0j} + \sum_k \beta_{jk} X_{jk}^h + \gamma D_j^h$$

où V_j^h est l'utilité stricte pour le type de logement j définie selon le mode d'occupation, la grandeur du logement et la densité; α_{0j} est une constante spécifique à l'alternative j; X_{jk}^h représente la kème variable spécifique de telle sorte que son coefficient β_{jk} varie

selon les alternatives et D_j est le prix minimal de l'alternative j défini comme une variable générique de telle sorte que son coefficient γ ne varie pas avec les alternatives.

La constante spécifique α_{0j} est active pour la jème alternative et vaut 0 autrement. On ne définit seulement que J-1 constantes spécifiques afin d'éviter le problème de multicollinéarité parfaite. Le rôle de ces constantes s'apparente à celui qu'elles jouent dans la régression linéaire. Elles permettent de capter l'influence moyenne des variables exclues de la fonction d'utilité. De plus, dans le cas d'un échantillon non aléatoire comme celui dont on dispose pour la présente étude, les constantes assurent que les coefficients estimés des attributs inclus de la fonction d'utilité soient non biaisés. Les estimateurs de ces constantes sont alors biaisés mais il est possible de corriger pour la présence d'un tel biais en ajustant la valeur de ces constantes spécifiques (Anas (2)). Dans le cas simple de deux alternatives (j=2), la fonction d'utilité stricte de chaque alternative peut alors s'écrire:

$$40) \quad V_1^h = \alpha_1 + \sum_k \beta_{1k} X_{1k}^h + \gamma D_1^h \quad \text{et} \quad V_2 = \gamma D_2^h.$$

On constate que les variables socio-économiques n'apparaissent pas dans l'alternative de référence puisque ces variables sont définies comme des variables spécifiques aux alternatives.

2.2 L'estimation du modèle logit multinomial

Le modèle de choix probabiliste peut être estimé en considérant simultanément l'ensemble des choix ou encore en exploitant

l'hypothèse d'indépendance des alternatives sur lequel repose le modèle. On estime alors de façon séquentielle en décomposant les choix individuels. Cette dernière procédure nous a fourni de meilleurs résultats et c'est celle que nous reproduisons ici.

Lorsque la décision est complexe, comme c'est le cas pour le choix d'un logement, il n'est pas inconcevable de penser que l'individu décompose le processus décisionnel et suit un cheminement arborescent. Il se comporte comme si l'utilité qu'il retire d'un bien peut être décomposée au niveau des caractéristiques qui le composent de telle sorte qu'il n'a besoin de considérer à chaque étape qu'un nombre limité d'alternatives. La fonction d'utilité pour le type de logement est séparable et additive et elle peut être exprimée de la façon suivante:

$$41) \quad U_j^h = U_{dg\ell}^h = V_\ell^h + V_{g\ell}^h + V_{dg\ell}^h + E_{dg\ell}^h$$

où $U_{dg\ell}^h$ est la fonction d'utilité d'un ménage h associé au choix du logement de densité d, de grandeur g et de mode d'occupation ℓ. Les probabilités conditionnelles sont données par:

$$42) \quad P_\ell^h = P[V_\ell \geq V_{\ell'}; \qquad \text{pour tout } \ell' \neq \ell]$$

$$43) \quad P_{g|\ell}^h = P[V_{g\ell} \geq V_{g'\ell}; \qquad \text{pour tout } g' \neq g]$$

$$44) \quad P_{d|g\ell}^h = P[V_{dg\ell} \geq V_{d'g\ell}; \qquad \text{pour tout } d' \neq d]$$

A partir des équations précédentes, on peut obtenir la probabilité conjointe, à savoir:

45) $$P^h_j = P^h_{dg\ell} = P^h_\ell \cdot P^h_{g|\ell} \cdot P^d_{d|g\ell}$$

Une telle décomposition du processus de décision simplifie énormément le problème d'estimation car au lieu d'estimer un modèle sur les 12 alternatives possibles, on a besoin de considérer au plus 3 alternatives par étape. Pour illustrer davantage la décomposition retenue, considérons l'éventualité suivante: si vous décidez d'acheter une propriété de grande dimension, quelle densité choisirez-vous? Notre hypothèse prévoit que la réponse à cette question dépend uniquement des attributs des diverses alternatives de densité et est indépendante des autres attributs spécifiques à la grandeur du logement ou à son mode d'occupation. Pour qu'une telle hypothèse soit acceptable, il faut que la fonction d'utilité soit additive et séparable. On doit supposer en fait que les taux marginaux de substitution entre les attributs propres à la densité sont constants quelle que soit la grandeur du logement ou le choix du mode d'occupation. Une telle hypothèse apparaît assez forte dans le cas du logement où effectivement la substitution entre différents niveaux de densité dépend de la dimension du logement ou encore du mode d'occupation. Les modèles logit permettent de vérifier empiriquement la validité d'une telle hypothèse. Ce sont les modèles séquentiels avec valeur inclusive, McFadden (16).

L'estimation de ce modèle hiérarchique a été effectuée en décomposant de façon appropriée notre échantillon de 1,351 ménages. En se référant à la figure 2, les estimations aux deux premiers niveaux portent sur un modèle à deux alternatives alors qu'au troisième niveau, elles portent sur un modèle de choix multiples. Les variables

utilisées sont d'une part les variables socio-économiques telles l'âge du chef, la taille du ménage, le revenu et d'autre part, les prix minimaux des diverses alternatives telles que calculés à la section précédente. Comme nous l'avons déjà mentionné, les variables socio-économiques sont introduites comme des variables spécifiques aux alternatives et les prix sont des variables génériques. Ainsi, les coefficients des variables socio-économiques varient selon les alternatives et ceux des prix sont constants d'une alternative à l'autre. Les résultats des estimations sont présentés au tableau 2 où les équations se lisent verticalement.

En examinant ces résultats, on constate que la plupart des coefficients sont significatifs et du bon signe. Les $(rho)^2$ qui servent à mesurer la qualité de l'ajustement statistique ne sont pas très élevés sauf pour les équations (1), (3) et (6) où leur valeur indique que la capacité explicative des variables retenues est relativement bonne. Pour fins d'illustration, examinons la première équation sur le choix du mode d'occupation. Pour mieux comprendre la sensibilité des choix aux variations dans les variables explicatives, il convient d'examiner au tableau 3 les élasticités directes et les élasticités croisées du choix du mode d'occupation en fonction des variables socio-économiques et des prix de chaque mode.

TABLEAU 2

RESULTATS DE REGRESSION DU MODELE HIERARCHIQUE

	Niveau 1 Choix du mode d'occupation	Niveau 2 Choix de la grandeur du logement		Niveau 3 Choix du type de logement			
	P[P] (1)	P[P|P] (2)	P[P|L] (3)	P[DE|P,P] (4)	Pr[DE|P,G] (5)	Pr[PX|L,P] (6)	Pr[PX|L,G] (7)
Constantes spécifiques	- 3,340 (4.02)	3,148 (5.99)	2,220 (5.64)	-	-	-	-
- détaché/plex	-	-	-	2.99 (3.83)	2.71 (2.54)	-1.29 (1.62)	6.62 (2.37)
- attaché/walk-up	-	-	-	2.49 (2.44)	1.13 (0.98)	0.25 (0.37)	5.91 (2.09)
Age	0.021 (4.48)	-0.026 (3.73)	-0.005 (0.815)	-	-	-	-
- détaché/plex	-	-	-	-0.041 (3.80)	-0.031 (1.87)	-0.037 (4.14)	-0.087 (2.45)
- attaché/walk-up	-	-	-	-0.046 (3.01)	-0.025 (1.42)	-0.045 (5.61)	-0.087 (2.59)
Taille	0.736 (13.17)	-0.155 (2.57)	-0.618 (8.04)	-	-	-	-
- détaché/plex	-	-	-	-0.254 (2.59)	-0.089 (0.79)	0.769 (4.11)	0.273 (0.82)
- attaché/walk-up	-	-	-	-0.209 (1.61)	-0.071 (0.53)	0.677 (3.81)	0.328 (0.97)
Revenu (000)	0.106 (11.69)	-0.039 (4.44)	0.003 (0.03)	-	-	-	-
- détaché/plex	-	-	-	0.044 (2.50)	0.019 (1.02)	-0.071 (3.61)	-0.160 (3.39)
- attaché/walk-up	-	-	-	0.025 (1.20)	0.019 (0.93)	-0.061 (3.61)	-0.160 (3.28)
Loyer (00)	-0.712 (1.90)	0.279 (0.75)	-1.012 (2.43)	-0.959 (4.93)	-1.41 (4.19)	-2.23 (5.41)	-1.099 (1.96)
* observations	1,351	735	616	449	286	410	206
$(rho)^2$	0.25	0.04	0.10	0.08	0.06	0.12	0.08
χ^2	477.2	36.1	80.7	60.9	32.5	104.3	27.3

Le χ^2 est définie comme $-2\,[\log L^*(0) - \log L^*(\hat{\beta})]$ où L^* est la valeur maximum de la fonction de vraisemblance sous l'hypothèse que tous les paramètres sont nuls $L^*(0)$ et sous l'hypothèse que les paramètres prennent les valeurs estimés $L^*(\hat{\beta})$.

Le $(rho)^2$ est définie comme $1 - [L^*(\hat{\beta})/L^*(0)]$

Tableau 3

ELASTICITES DIRECTES ET CROISEES
CHOIX DU MODE D'OCCUPATION

	P(P) Elasticités directes	P(L) Elasticités croisées
Age du chef	.2654	- .5737
Taille du ménage	.6898	-2.0196
Revenu du chef	.6182	-1.8125
Prix du logement en propriété	-1.1594	2.6978
Prix du logement locatif	1.1772	- .8024

Comment interpréter ces résultats? D'abord, l'élasticité directe mesure l'impact d'un accroissement de 1 pour cent dans la valeur d'une variable explicative sur la probabilité d'acheter son logement. Par exemple, l'accroissement du revenu de 1 pour cent augmente la probabilité d'être propriétaire de 0,6182 pour cent, les autres variables demeurant constantes. L'élasticité croisée par contre indique que ce même accroissement de revenu diminue la probabilité d'être locataire de 1,81 pour cent. La variable qui diminue le plus la probabilité d'être locataire est la taille du ménage. La probabilité d'être propriétaire est plus sensible à la variation du prix du logement que ce n'est le cas de la probabilité d'être locataire. La variable qui affecte le plus la probabilité d'être locataire est le prix du logement en propriété. En effet, l'élasticité croisée nous indique que l'augmentation de 1 pour cent du prix de la propriété entraîne une augmentation de 2,69 pour cent dans la probabilité d'être

locataire. Ce résultat est très important dans la mesure où tout programme qui a pour conséquence de diminuer le coût de l'accès à la propriété aura un effet considérable sur le locatif. On pourrait faire une analyse analogue dans le cas du choix de la grandeur du logement et de la densité.

A partir des équations estimées du tableau 2 et des probabilités conditionnelles calculées comme à l'équation 3), on peut reproduire les probabilités individuelles en se servant de l'équation 45). Toutefois, pour fins de simulation, il faut agréger ces probabilités par classes homogènes de ménages.

2.3 Le regroupement des ménages et les probabilités agrégées ex-ante

Pour le regroupement des ménages par classes homogènes, nous avons retenu trois catégories de revenu, trois tailles de ménage et trois catégories d'âge du chef, ce qui donne vingt-sept classes différentes. Ces catégories socio-économiques de même que les moyennes des variables dans l'échantillon sont reproduites au tableau 4. A partir des équations estimées du tableau 2 et des valeurs moyennes des variables socio-économiques, nous avons calculé les probabilités conditionnelles associées au choix du logement selon la densité, la grandeur et le mode d'occupation. En se servant de l'équation 45), nous calculons les probabilités agrégées ex-ante P_{ij} selon la classe de ménage i (contenant N_i ménages) et le type de logement j. Nous obtenons ainsi un tableau de probabilités de choix de dimension I x J, soit 27 x 12; cependant, ces probabilités ne respectent pas exactement l'équation 12) sur le nombre M_j d'unités du type de logement j choisies dans l'échantillon.

Tableau 4

CATEGORIES SOCIO-ECONOMIQUES
ET MOYENNES DES VARIABLES

	Nombre de ménages[a]	Moyennes des variables
Revenu:		
0 - $14 999	557	$ 9 260
$15 000 - $24 999	536	$19 700
$25 000 et plus	258	$32 500
Taille du ménage		
1 personne	202	1
2 personnes	401	2
3 et plus	748	4,02
Age du chef		
Moins de 35 ans	663	27,5
35 - 55 ans	466	44,6
55 ans et plus	222	66,2

2.4 Les probabilités ajustées et les probabilités ex-post

Afin d'obtenir les probabilités ajustées qui respectent les contraintes de stock M_j, nous avons résolu le problème présenté aux équations 28)-30). Cet ensemble de probabilités constitue le point de référence de l'analyse des scénarios, ce qu'on appellera

le scénario de référence. Pour obtenir les probabilités de choix ex-post Q_{ij} selon les divers scénarios modifiant la composition N_i des classes de ménages et/ou les stocks disponibles S_j, nous avons résolu le problème 21)-23).

Ces deux modèles sont de la forme générale suivante:

46) $$\underset{x_j}{\text{Min}} \sum_{j=1}^{n} x_j \log\left(\frac{x_i}{y_j}\right)$$

47) $$\text{sujet à } \sum_{j=1}^{n} a_{ij}\, x_j = b_i \ (\text{ou} \geq b_i) \qquad (i=1,\ldots,m)$$

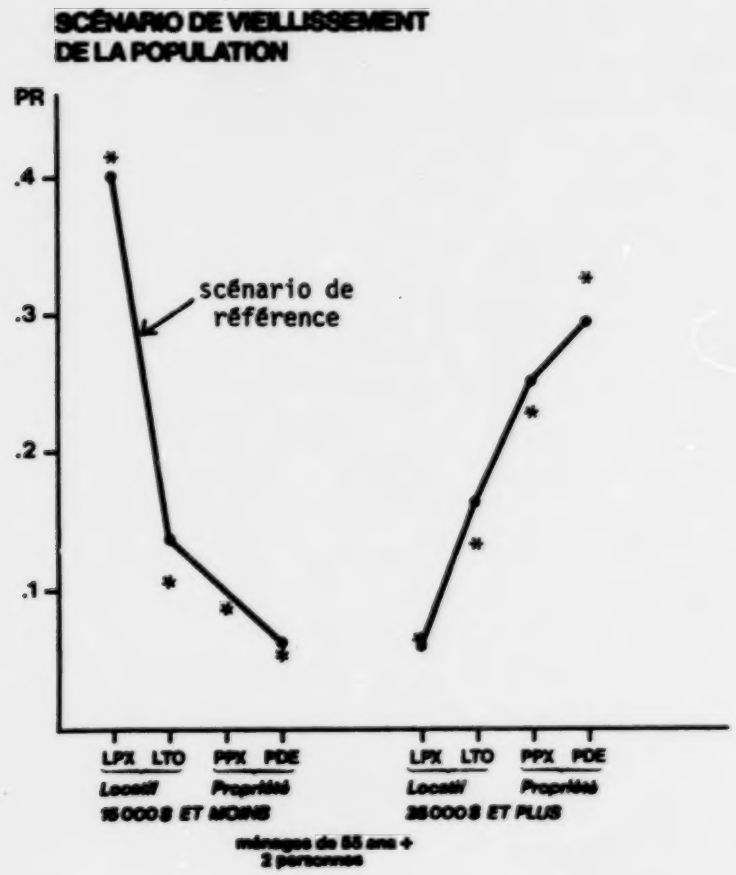
où $b_i \geq 0$, $a_{ij} \geq 0$ et $y_i \geq 0$ sont des constantes. Pour les deux problèmes précédents, les coefficients a_{ij} sont 1 ou 0 (si le type de logement j n'est pas disponible pour la classe i) et les contraintes décrivant la structure d'un problème de transport. Ces problèmes non linéaires ont d'abord été résolus par la méthode de décomposition de Frank-Wolfe, puis de façon beaucoup plus efficace par la technique MART (multiplicative algebric reconstruction method) telle que décrite par Lent (13).

3. ANALYSE DE QUELQUES SCENARIOS

Plusieurs scénarios ont été analysés. Pour fins de concision, nous n'en présentons que deux; un premier sur l'impact du vieillissement de la population et un second sur l'impact de la diminution de la taille des ménages. Pour ces deux scénarios, la composition du stock de logements disponibles n'a pas été modifiée, c'est-à-dire que $S_j = M_j$, pour tout j. Le scénario de vieillissement de la population repose sur l'hypothèse que la proportion du nombre de ménages dans la classe des moins de 35 ans passe de 45% qu'elle était dans le scénario de référence à 35%, le différentiel étant reporté de façon proportionnelle sur les autres classes. Pour mieux visualiser l'impact d'un tel scénario sur les probabilités de choix, nous nous limitons au groupe de ménages de 55 ans et plus et de 2 personnes. La figure 3 reproduit les probabilités de référence et les nouvelles probabilités.

On constate d'une part que pour les faibles revenus soit, $15 000 et moins, la probabilité de choisir le locatif est beaucoup plus élevée que celle de choisir la propriété et l'inverse est vrai pour les revenus élevés de $25 000 et plus. Suite au vieillissement de la population, il s'exerce une pression sur les logements plus denses tant sur le locatif que sur la propriété et entraîne des substitutions. Chez les forts revenus, la substitution va se faire en faveur des maisons unifamiliales détachées, chez les faibles revenus en faveur des plex locatifs. Ceci est la résultante de la modification des prix relatifs. Par exemple, on aurait pu prévoir à l'aide d'un tel modèle que

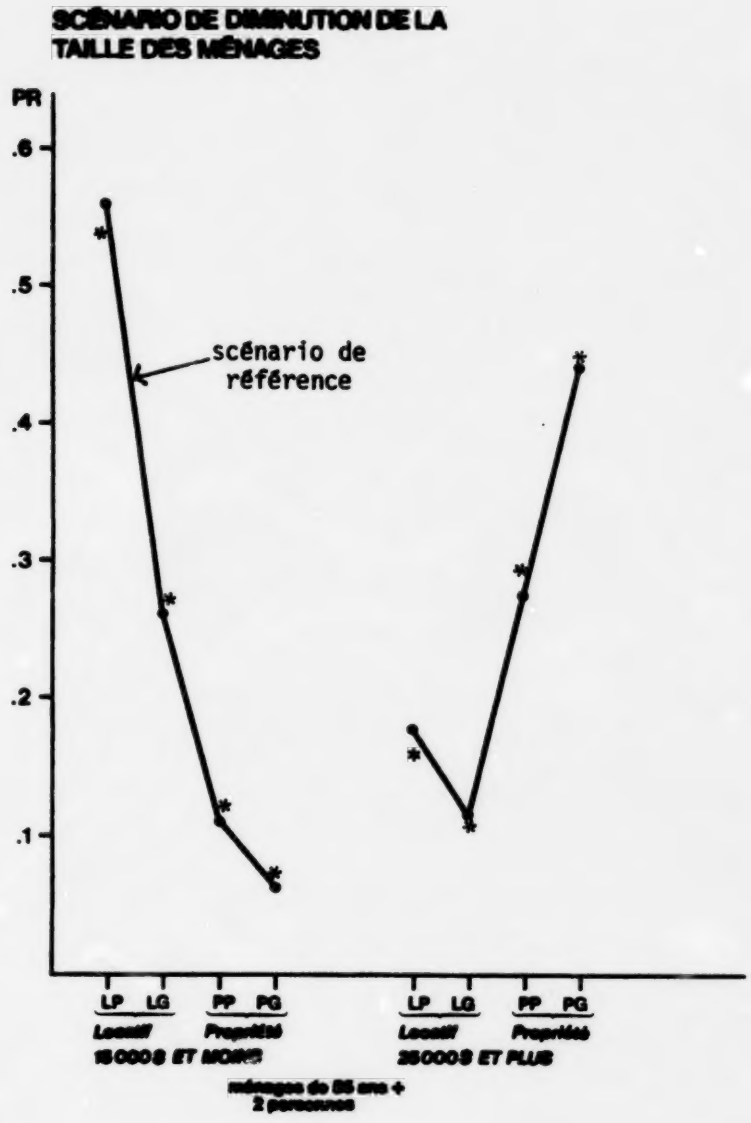
Figure 3



le vieillissement de la population offrait un marché intéressant pour les unités plus denses détenues en propriété: c'est la formule du condominium.

Le second scénario consiste en une diminution de la taille des ménages. On fait l'hypothèse dans ce scénario que la proportion des ménages d'une seule personne passe de 15% qu'elle était dans le scénario de référence à 20%. L'impact d'un tel changement sur les probabilités de choix est présenté à la figure 4. On peut constater que pour les revenus

Figure 4



élevés, il y a substitution en faveur de la propriété mais surtout de petits logements. Chez les faibles revenus, il y a une substitution en faveur des logements en propriété et des grands logements locatifs. La pression qui s'exerce alors sur le locatif entraîne une modification des prix relatifs en faveur de la propriété.

4. CONCLUSION

Le modèle de simulation proposée présente plusieurs aspects nouveaux sur la modélisation des choix résidentiels. L'utilisation des modèles de choix probabilistes permet d'allier les fondements théoriques des comportements individuels avec la partie aléatoire de ces comportements lorsqu'on veut les analyser de façon empirique. Le modèle combine également ce premier type d'approche basé sur les comportements avec un modèle de programmation mathématique déterministe assurant que les contraintes sur les disponibilités de l'offre soient respectées. Ces modèles sont compatibles du point de vue formel et offrent une grande souplesse à l'étape de la simulation. En effet, ils permettent d'évaluer l'impact de simulations tant du côté de la demande que de l'offre.

Les deux simulations présentées dans le texte ne sont qu'un pâle reflet des possibilibés d'un tel modèle. Il est possible en effet de simuler l'impact sur les prix relatifs et les probabilités de choix de chaque alternatives d'une variation dans la composition du stock. Il est également possible, en utilisant l'estimation du modèle logit multinomial, de saisir l'impact sur les probabilités de référence d'une augmentation des coûts de transport ou encore de la croissance des revenus.

BIBLIOGRAPHIE

(1) ANAS, A., Residential Location Markets and Urban Transportation, Academic Press, 1982.

(2) ANAS, A., «Discrete Choice Theory, Information Theory and Multinomial Logit and Gravity Models», Transportation Research, Vol. 17, Nº 1, 1983.

(3) Association Canadienne des Automobilistes, «Les dépenses d'automobile», 1980.

(4) CENSOR, Y. and G.T. HERMAN, «An Iterative Row-Generation Methods for Feasibility and Optimization Problems Involving Sparse Matrices and Their Applications», Technical Report M1PG18, Department of Computer Science, SUNY at Buffalo, Amherst, N.Y., 1978.

(5) Conseil des transports de la région de Montréal (COTREM), «Recodage de l'enquête origines-destinations en un système de 200 zones», Communauté Urbaine de Montréal, juin 1980.

(6) DANSEREAU, F. et G. FORTIN, «Les NER de Montréal et de Québec: traits généraux de l'univers et de l'échantillon», INRS-URBANISATION, Etudes et Documents, Nº 15.

(7) DE LEEUW, F. et R.J. STRUYK, «The Web of Urban Housing: Analysing Policy with a Simulation Model», The Urban Institute, Washington, D.C., 1975.

(8) DESROSIERS, J. et A. LAPOINTE, «Evaluation de l'impact des changements structurels sur les choix de résidence», H.E.C., GERAD, 1982.

(9) DOMENCICH, T. et D. McFADDEN, Urban Travel Demand: A Behavioral Analysis, North-Holland Pub., 1975.

(10) ELFVING, T., «On Some Methods for Entropy Maximization and Matrix Sealing», Linear Algebra and Applications, Vol. 34, décembre 1980, pp. 321-339.

(11) HOBSON, A., «A New Theorem of Information Theory», Journal of Statistical Physics, 1969.

(12) INGRAM, G., J.F. KAIN et al., The National Bureau of Economic Research, Detroit Prototype, Urban Simulation Model, Columbia University Press, 1972.

(13) LENT, A., «A Convergent Algorithm for Maximum Entropy Image Restoration, With a Medical X-Ray Application», Image Analysis and Evaluation, R. Shaw, editor, pp. 249-257, SPE Conference Proceedings, Toronto, Canada, 1976.

(14) LOS, M., «Discrete Choice Modelling and Desequilibrium in Land and Transportation Planning», Working Paper N° 137, Centre de Recherche sur les Transports, Université de Montréal, 1979.

(15) McFADDEN, D., «Conditional Logit Analysis of Qualitative Choice Behavior», Frontiers in Econometrics, Zarembka, editor, Academic Press, N.Y., 1974.

(16) McFADDEN, D., «Modelling the Choice of Residential Location», Spatial Interaction and Planning Theory, Kaklquist et al., 1978.

(17) QUIGLEY, J., «Housing Demand in the Short Run: An Analysis of Polytomous Choice», Explorations in Economic Research, 3, 1976.

(18) RENYI, A., Calcul des probabilités, Dunod, Paris, pp. 496-510, 1957.

(19) SNICKARS, F. et J.W. WEIBULL, «Information Principle: Theory and Practice», Regional Science and Urban Economics, 7, 1977.

LISTE DES CAHIERS DE RECHERCHE DEJA PUBLIES PAR LES PROFESSEURS DE L'INSTITUT D'ECONOMIE APPLIQUEE

No	Auteur(s)	Titre
81-11	BRENNER, G.A. BRENNER, R.	Why Do People Gamble?
81-14	KOSTECKI, M.M. SECK, D.	Treasury Revenue and Foreign Trade Taxation.
81-16	BRENNER, G.A. BRENNER, R.	The Easy Case for Progressive Taxation.
81-22	MARCHON, M.N. VanPEETERSSEN, A.	Impact de la substituabilité entre facteurs causé par le changement du prix relatif de l'énergie.
81-23	LAPOINTE, A.	La télévision à péage: bilan des aspects technico-économiques.
82-09	BRENNER, G.A. BRENNER, R.	Why Do Productivity and Profits Decline? Theory and Evidence.
82-10	BRENNER, G.A. BRENNER, R.	Les loteries - Pourquoi les Québécois y participent-ils?
82-15	MARCHON, M.N. OUELLETTE, P.	Comptabilité nationale du secteur privé au Canada.
82-20	LAPOINTE, A. MOISAN, H.	Elasticité de la demande de logement au Canada: nouveaux résultats sur des données désagrégées.
82-21	DAGENAIS, D.L.	L'utilisation d'un modèle pour fins d'études de validation de mécanismes de sélection.
82-22	BRENNER, G.A.	What Should the Optimal Bufferstocks of Grains Be?
82-23	BRENNER, G.A.	Why Did Inheritance Laws Change?
82-24	GAUTHIER, G. MOISAN, H.	La subvention à la garderie: typologie et analyse.
82-25	GAUTHIER, G. MOISAN, H.	Analyse des coûts dans une garderie au Québec pour chacun des postes budgétaires.

82-29	KOSTECKI, M.M.	The Consumer in a Socialist Economy.
82-30	KOSTECKI, M.M.	Note on the Impact of Government Takeovers in Mineral Industries on International Markets.
83-03	LeGOFF, J.P.	La notion de stratégie industrielle.
83-06	BRENNER, G.A.	Why Do People Gamble? Further Canadian Evidence.
83-14	LEROUX, F. ZACCOUR, G.	Structures de marché et intensité de la publicité.
83-16	RAYNAULD, J. TESSIER, J.	Risk Premiums in Future Markets: An Empirical Investigation.
83-17	GREGORY, A.W. RAYNAULD, J.	An Econometric Model of Canadian Monetary Policy Over the 1970's.
83-23	BRENNER, G.A. BRENNER, R.	Who are the Entrepreneurs? (Or, Don't Confuse Brains with a Bull Market).
84-06	BRENNER, R. COURVILLE, L.	Fraud Among Buyers and Sellers: Social, Legal and Market Remedies.
G-13-19	DESROSIERS, J. LAPOINTE, A.	Evaluation de l'impact des changements structurels sur des choix de résidence.
IEA-84-01	LAPOINTE, A. DESROSIERS, J.	Modelisation des choix de résidences.
IEA-84-02	NAPPI, C.	Structural Changes in the International Aluminum Industry: The Situation in Canada.